

Research Data Management

Resources from the Digital Curation Centre

Alex Ball

17 September 2014

Hello, my name is Alex Ball and I work for the Digital Curation Centre (DCC). In case you haven't heard of us, ¶ we are a collaboration between

- University of Edinburgh
- HATII, University of Glasgow
- UKOLN, University of Bath, where I am based.

For over ten years we've identified, compiled and disseminated best practice in digital curation in various ways, and we've developed some useful tools supporting research data management. In the course of this presentation I'll give you an overview of the resources we have to offer: ¶

Contents

1. Introduction.....	2
2. Preparing for RDM	2
3. Requirements Analysis	4
3.1. Data Asset Framework.....	4
3.2. CARDIO.....	5
4. Policy and strategy	6
5. Data management planning	6
5.1. DMPonline.....	7
6. Managing active data	8
7. Data selection and handover.....	9
8. Data repositories and catalogues	10
9. Guidance, training and support.....	11

1 Introduction

Until very recently, institutions (in the UK at least) did not have to care very much about research data management. The disciplines that really cared about data sharing – typically those making unrepeatable observations or using breathtakingly expensive equipment – had made their own arrangements with dedicated data centres and repositories. And it was these centres that provided advice and support to researchers.

But when the UK signed up to the OECD *Declaration on Access to Data from Public Funding* in 2004 it pushed the issue of open data up the political agenda, and research funders started looking harder at what was happening to the data they had paid for. Meanwhile, the closure of the Arts and Humanities Data Service in 2008 and the ‘climategate’ scandal of 2009–10 were warnings for institutions to take data management more seriously. The game changer came in 2011, though, when the EPSRC published their data policy. ¶ The other major funders laid funding conditions on applicants. The EPSRC laid them on institutions:

- EPSRC expects all those [research organisations] it funds to have developed a *clear roadmap* to align their policies and processes with EPSRC’s expectations by 1st May 2012, and to be *fully compliant* with these expectations by 1st May 2015. •

Here are EPSRC’s expectations (paraphrased):

1. *Research organisations (ROs)* to raise awareness of data sharing responsibilities and issues.
2. Publications should link to underlying data.
3. ROs must keep track of their research datasets and requests for them.
4. Born-analogue data must also be shareable on request.
5. ROs must provide open, online catalogues of their data; digital data must be given a robust ID.
6. Access restrictions should be clear and justified.
7. ROs must provide access to data for 10 years from last access.
8. ROs must curate their research data.
9. ROs must pay for this from their existing public funding streams.

You can see just how much is laid here at the door of institutions. So this will be a strong driver for any institution with significant EPSRC funding to get involved with research data management, but it’s not the only one. ¶

2 Preparing for RDM

If you need convincing, or need to convince others at your institution, a great place to start is with our briefing paper *Making the Case for Research Data Management* (Figure 1).

- Drivers
- Building the Services
- Identifying Benefits and Challenges
- Creating the Environment
- Looking to the Future

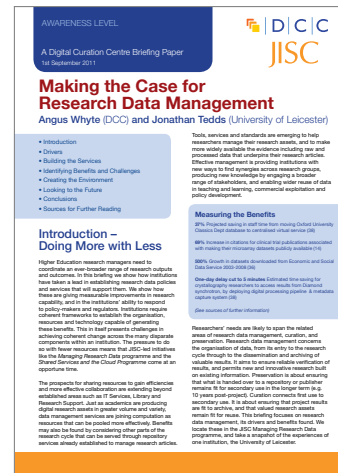


Figure 1: *Making the Case for Research Data Management* (see <http://www.dcc.ac.uk/resources/briefing-papers/making-case-rdm>)

1. Take stock
2. Let research needs drive your strategy
3. Re-evaluate your existing infrastructure and data architecture
4. Get to know the new technologies and standards
5. Bring your staff up to speed

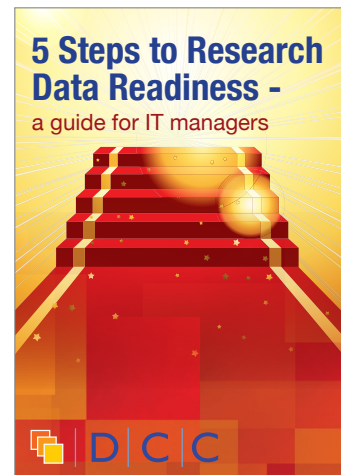


Figure 2: *5 Steps to Research Data Readiness* (see <http://www.dcc.ac.uk/resources/briefing-papers/five-steps-research-data-readiness>)

It sets out very clearly the drivers for and benefits of RDM for institutions, and includes a case study of how the challenges were tackled at the University of Leicester. ¶

So, having established that you need to develop RDM support at your institution, how do you go about it? If you want the short answer we have a handy little booklet called *5 Steps to Research Data Readiness* (Figure 2). It only takes a few moments to read, but in that time you'll get the basic idea of what's involved. ¶

For a more comprehensive answer, the resource you need is our guide *How to Develop Research Data Management Services* (Figure 3). As the title suggests, it outlines a process for developing RDM services within your institution, outlines the components you will need to consider, and gives specific advice on setting up each one in turn. Quite often, it points you in the direction of further resources which are specific to a stage or component, so let's have a look at some of those provided by the DCC. ¶

1. Define scope, define teams
2. Requirements analysis
3. Pilot services, implementation

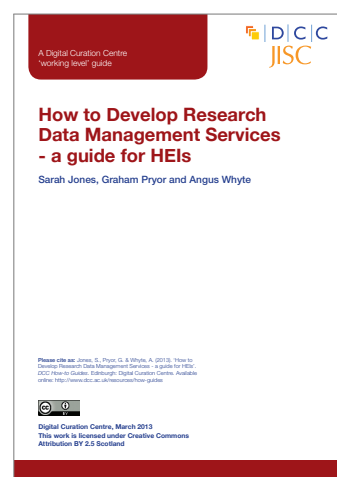


Figure 3: *How to Develop Research Data Management Services* (see <http://www.dcc.ac.uk/resources/developing-rdm-services>)

3 Requirements Analysis

- Data Management Roles and Responsibilities
- Research Data Management in Context
- Development phases
- Getting Started and Discovering Requirements
- Next Steps and Future Challenges

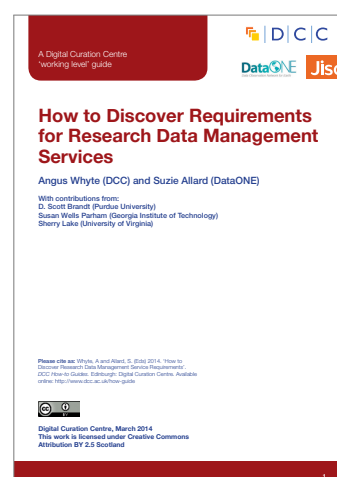


Figure 4: *How to Discover Requirements for Research Data Management Services* (see <http://www.dcc.ac.uk/how-discover-requirements>)

Before building a set of services it is important to establish what they should achieve. That includes getting an accurate picture of the current state of affairs, identifying areas of improvement and prioritising them. We have another guide, *How to Discover Requirements for Research Data Management Services*, that can help you through that process (Figure 4). But we have also developed two methodologies for getting to those concrete results. ¶

3.1 Data Asset Framework

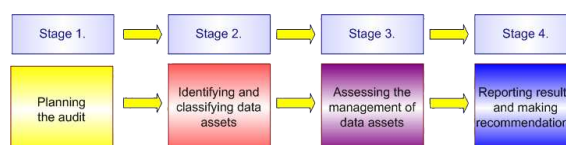


Figure 5: Stages of the DAF (see <http://data-audit.eu/>)

To get a handle on what data an institution holds, there's the Data Asset Framework, or DAF (Figure 5). This was originally intended as a methodology for

assembling an inventory of data assets, based on desk research, interviews and questionnaires. But what its users found really valuable were the insights it gave them into the state of current practice and the scale and variety of the data assets out there in the wild.

If you're in the early stages of setting up RDM support at your institution, this is a good place to start. The methodology itself is fully documented, and we also provide tips for implementing it in various situations, along with sample questionnaires and interview templates. ¶

3.2 CARDIO

Taking things one step further there's CARDIO, with which you are all now familiar. If you're thinking it might be useful to gauge your institution's readiness for research data management, you can use the same paper-based version we've used today¹ or adapt it to suit your own circumstances, but the DCC provides two online tools you can use off the shelf.

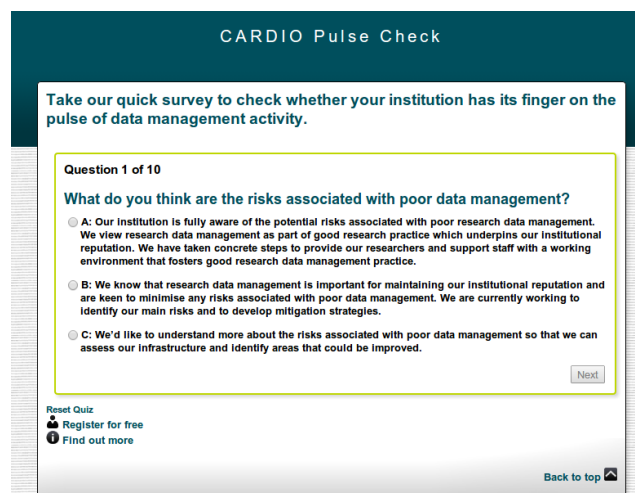
The image shows a screenshot of the 'CARDIO Pulse Check' online survey. The header is a dark teal bar with the text 'CARDIO Pulse Check' in white. Below the header, a light blue box contains the text: 'Take our quick survey to check whether your institution has its finger on the pulse of data management activity.' The main content area is white and contains 'Question 1 of 10'. The question is 'What do you think are the risks associated with poor data management?'. There are three radio button options: A: 'Our institution is fully aware of the potential risks associated with poor research data management. We view research data management as part of good research practice which underpins our institutional reputation. We have taken concrete steps to provide our researchers and support staff with a working environment that fosters good research data management practice.' B: 'We know that research data management is important for maintaining our institutional reputation and are keen to minimise any risks associated with poor data management. We are currently working to identify our main risks and to develop mitigation strategies.' C: 'We'd like to understand more about the risks associated with poor data management so that we can assess our infrastructure and identify areas that could be improved.' A 'Next' button is at the bottom right of the question box. At the bottom left, there are links: 'Reset Quiz', 'Register for free', and 'Find out more'. At the bottom right, there is a 'Back to top' link with an upward arrow icon.

Figure 6: CARDIO Pulse Check (see <http://cardio.dcc.ac.uk/quiz>)

The first is very quick and easy (Figure 6): ten multiple-choice questions that guide you through the main areas of data management and invite you to reflect on how well you're doing in each of them. ¶

The second is rather more thorough (Table 1). It invites you and other stakeholders to assess the institution's performance in 30 different areas, and provides facilities for getting a consensus view and formulating a concrete action plan. This might include writing policies on IPR or risk management, rethinking how IT facilities are financed, providing new infrastructure such as a data catalogue or repository, or providing data management training. ¶

¹CARDIO, EPSRC version: http://www.dcc.ac.uk/webfm_send/1666 (PDF), http://www.dcc.ac.uk/webfm_send/1667 (MS Word).

Table 1: CARDIO Process (see <http://cardio.dcc.ac.uk/>)

Organisation	Technology	Resources
<ol style="list-style-type: none"> 1. Data Ownership and Management 2. Data Policies and Procedures 3. Data Policy Review 4. Sharing of Research Data/Access to Research Data 5. Preservation and Continuity of Research 6. Internal Audit of Research Activities 7. Monitoring and Feedback of Publication 8. Metadata Management 9. Legal Compliance 10. Intellectual Property Rights and Rights Management 11. Disaster Planning and Continuity of Research 	<ol style="list-style-type: none"> 1. Technological Infrastructure 2. Appropriate Technologies 3. Ensuring Availability 4. Managing data integrity 5. Obsolescence 6. Managing technological change 7. Security Provisions 8. Security Processes 9. Metadata tools 10. Institutional Repository 	<ol style="list-style-type: none"> 1. Data Management Costs and Sustainability 2. Business Planning 3. Technological Resources Allocation 4. Risk Management 5. Transparency of Resource Allocation 6. Sustainability of Funding for Data Management and Preservation 7. Data Management Skills 8. Number of Staff for Data Management 9. Staff Development Opportunities

4 Policy and strategy

Once you've gathered your requirements and established what needs to be done, you then need to put it into practice. If you are looking to develop an RDM policy at your institution we have plenty of guidance to help you. ¶

- **Funders' data policies**
 - **Overview:** A table and short summaries comparing research funders' policies.
 - **Detail:** Detailed overview of each funder's policy, stating requirement for data plans, expectations on data sharing and available support.
- *Five Steps to Developing a Research Data Management Policy*, similar to the other Five Steps leaflet I showed you before.
- **Research data policy briefing** to assist universities to create institutional RDM policies, with examples of different approaches.
- **Institutional data policies:** A table listing examples of UK universities' research data policies.
- **EPSRC Roadmaps:** A table listing examples of UK universities' roadmaps towards meeting EPSRC expectations on compliance with their Policy Framework on Research Data, with pointers to advice on preparing a roadmap.
- **Bibliography of further tools and resources**
- **Case study: RDM strategy: moving from plans to action**

Links to all these can be found in the Policy and Legal Resources section of our website (<http://www.dcc.ac.uk/resources/policy-and-legal>) or, in the case of the last, from <http://www.dcc.ac.uk/resources/developing-rdm-services>. ¶



5 Data management planning

Let's move on to the research lifecycle, which starts with data management planning.

Or at least it should do. Most UK funders require Data Management Plans, and with good reason. ¶ Writing and using a Data Management Plan helps

- to co-ordinate the actions of data stakeholders
- to ensure all necessary tasks are accomplished
- to ensure data are properly curated
- with releasing data in a timely fashion, and because DMPs encourage thought about metadata and consent forms right at the start,
- with sharing data as openly as possible
- with preserving data for future use

But they can be a bit tedious to write. After all, when they're writing their proposals researchers are brimming with big ideas, and they don't want to be bogged down in what must seem like strange administrative stuff, especially if the deadline is looming. ¶

5.1 DMPonline

To take the pain away, the DCC runs a service called DMPonline, which lets people:

1. create, store and update Data Management Plans; the templates can have multiple phases so researchers can add more detail at later stages of the research
2. meet both institutional and funders' data-related requirements, all in one go
3. receive specific guidance from funders and institutions
4. export Data Management Plans in various formats such as Word or PDF

We've already programmed in templates for all the major UK funders, with a wealth of guidance notes that researchers can read as they work through them. But we can't account for the particular circumstances in each institution; what we *can* do is give institutions the power to do that for themselves. ¶ Using our new admin interface (Figure 7), you can

- Add an extra section to funder templates
- Add whole additional templates; these can have any number of questions, arranged into any number of sections, and can have multiple phases.
- Add guidance text and links, either directly to particular questions, or indirectly via the themes with which the questions are tagged
- Suggest answers or give examples for researchers to adapt
- Monitor how many plans have been created by whom at your institution

In addition to that, the whole thing is open source (specifically, it uses the GNU Affero GPL licence), so anyone can set up their own instance and modify it so long as they share their changes. This is a good option if you want to apply your own branding and use your own URL. ¶

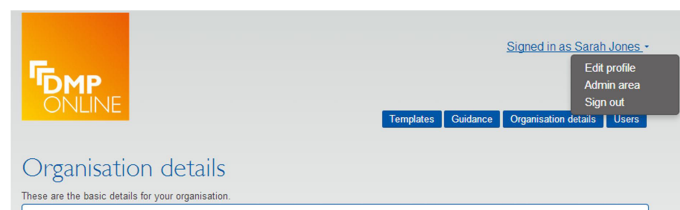


Figure 7: The DMPonline admin interface

6 Managing active data

Moving on to the active stage of research, the DCC has a range of resources intended to help researchers make their data as reusable, visible and preservable as possible. ¶

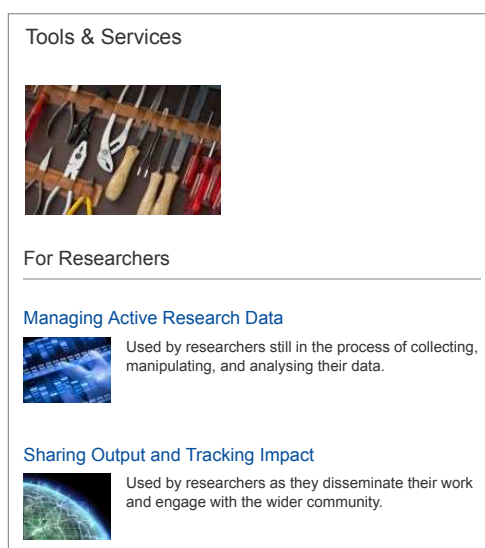


Figure 8: *Tools and Services Catalogue: For Researchers* (see <http://www.dcc.ac.uk/resources/external/tools-services>)

One of the first things they should look at is our Tools Catalogue (Figure 8). It provides detailed information on a range of tools aimed at researchers, from data storage and electronic lab notebooks through persistent ID management to impact tracking services. ¶

Speaking of impact, even if researchers make their data totally open and share it widely, it will be unusable to anyone else if it is not properly documented. But researchers are not always as familiar as you might hope with the documentation standards for their disciplines. For this reason, the DCC set up a catalogue of disciplinary metadata standards (Figure 9) where researchers or more likely support staff can go and look them up. This work has now gone international, with Metadata Standards Directory Working Group of the Research Data Alliance taking it on to bigger and better things. ¶

Another potential obstacle to reuse are intellectual property rights. The legal frameworks in place for protecting data are somewhat messy, and they are messy in different ways in different jurisdictions, so in this risk averse world researchers really do need to license their data appropriately if they want to see it reused. We have a guide for that

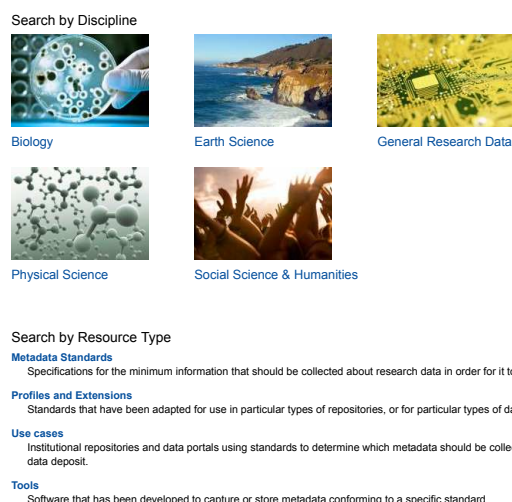


Figure 9: The DCC Disciplinary Metadata Catalogue (see <http://www.dcc.ac.uk/resources/metadata-standards>)

- Licensing concepts
- Prepared licences
- Bespoke licences
- Standard licences
- Multiple licensing
- Mechanisms for licensing data
- Licensing related information

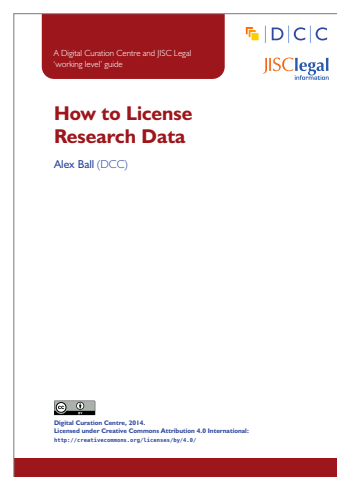


Figure 10: *How to License Research Data* (see <http://www.dcc.ac.uk/resources/how-guides/license-research-data>)

(Figure 10), recently revised, which explains the various licensing options and how one goes about applying a licence. ¶

A couple of years ago the Royal Society published a report, *Science as an Open Enterprise*, that argued that research data should be not only accepted but required as a part of the scientific literature. That implies that it should be subject to the same checks, balances and reward structures as traditional publications. To help that along (Figure 11), we have a guide to linking papers with their underlying data through the citation mechanism, and of course, citing third party data one has reused. ¶

7 Data selection and handover

As well as publishing what we might call the data behind the graph, researchers might have accumulated a lot of other useful data over the course of their project, and what

- Requirements for data citations
- Elements of a data citation
- Current issues and challenges
- Building a citation infrastructure
- Data citation infrastructures
- Current implementation issues

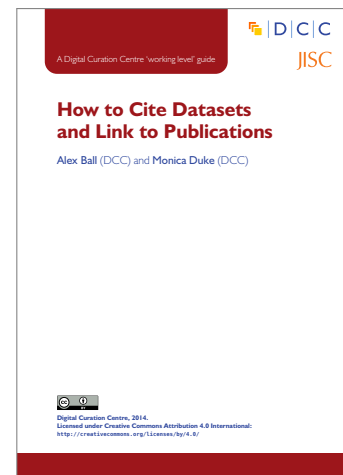


Figure 11: *How to Cite Datasets and Link to Publications* (see <http://www.dcc.ac.uk/resources/how-guides/cite-datasets>)

they ought to do is hand it over to a data centre or repository when the project ends. But repositories can't store everything, at least, not if they want to curate it properly. ¶

- Why select and appraise
- Appraisal Concepts
- Roles and Responsibilities
- Appraisal and Selection Policy
- Developing the Appraisal Process
- New Challenges and Opportunities

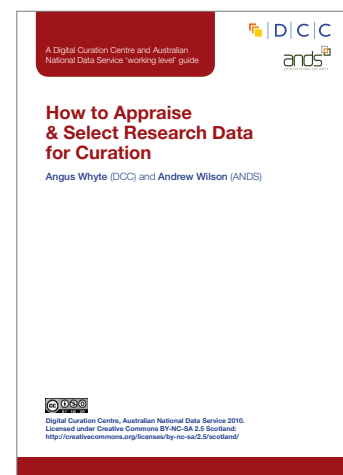


Figure 12: *How to Appraise and Select Research Data for Curation* (see <http://www.dcc.ac.uk/resources/how-guides/appraise-select-data>)

So to help repositories sort the proverbial wheat from the chaff, we have published the guide *How to Appraise and Select Research Data for Curation* (Figure 12). It explains how to develop a policy and process for appraisal and selection, so one can achieve a consistent collection and manage user expectations appropriately. ¶

8 Data repositories and catalogues

That's just one thing to bear in mind when setting up a data repository, but there are many more. ¶

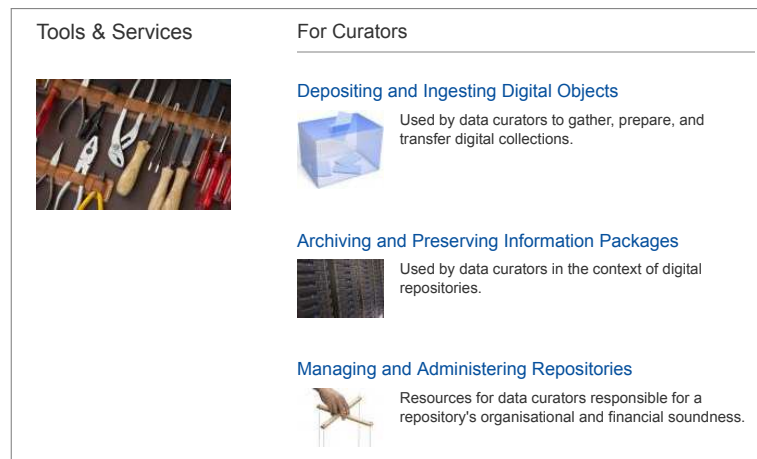


Figure 13: *Tools and Services Catalogue: For Researchers* (see <http://www.dcc.ac.uk/resources/external/tools-services>)

We have catalogued another set of tools, this time aimed at repository managers and other digital curators (Figure 13). The sorts of activities we cover include extracting metadata, migrating file formats, archival storage, providing access, writing policies and costing the service. ¶

- Improving Research Visibility – Getting Data on the Institutional Repository RADAR
- Storing and sharing data in an institutional repository – Hydra@Hull
- Assigning Digital Object Identifiers to Research Data at the University of Bristol



Figure 14: *Case studies relevant to data catalogues and repositories* (see <http://www.dcc.ac.uk/resources/developing-rdm-services>)

If you are curious what other institutions have done, and want to benefit from their experience, we’ve published some case studies that will interest you (Figure 14). They each concentrate on a different theme, so the case study of our hosts Oxford Brookes concerns the public catalogue side of things, while the one of Bristol focuses on assigning DOIs to datasets. ¶

9 Guidance, training and support

We’ve covered organisational and technical infrastructure, but there’s one component left we need to talk about and that’s the human infrastructure. It’s wonderful having

policies, procedures and tools in place, but they are no good if no-one knows about them or how to use them. That's where training comes in. ¶

- RDM Training for Librarians
- Increasing Participation in Internal RDM Training Sessions



Figure 15: Case studies relevant to research data management training (see <http://www.dcc.ac.uk/resources/developing-rdm-services>)

We have another couple of case studies related to training you could read (Figure 15): one on training librarians to support researchers in their data management, and one on training researchers directly. ¶

But we don't just talk about training, we do perform it ourselves.

- **Digital Curation 101:** an introduction to research data management and curation, the range of activities and roles that should be considered when planning and implementing new projects, and an overview of tools that can assist with curation activities – **materials available online**
- **Tools of the Trade:** how to use the suite of DCC tools.
- **Train the Trainer:** aimed at research support staff, librarians, and any other staff who need to train researchers in managing their data.
- **Curation webinars:** the first one was on the DMPonline admin interface; more are planned.
- If that is not enough, we also maintain a **Directory of third party RDM training materials** that you might borrow, or from which you might draw inspiration.

You can find all this at <http://www.dcc.ac.uk/training/> ¶

That, believe it or not, is just a selection of what we have to offer. I've concentrated on our practical resources, but we have a wealth of more theoretical and cutting-edge material if you want to go further:

- **Curation Reference Manual** provides deep insight and critiques on a range of curation topics;
- **International Journal of Digital Curation** showcases the latest research and developments in the area; many of the papers derive from material first presented at...
- **International Digital Curation Conference**, which is held annually. The next one will be held in London in February. The call for papers is out now, so do have a look and it would be great to see you there.



Alex Ball. DCC/UKOLN, University of Bath. <http://alexball.me.uk/>



Except where otherwise stated, this work is licensed under the Creative Commons Attribution 4.0 International licence: <http://creativecommons.org/licenses/by/4.0/>



The DCC is supported by Jisc.

For more information, please visit <http://www.dcc.ac.uk/>